A New Measure of Watermarking Security: The Effective Key Length

Patrick Bas, Teddy Furon

A New Measure of Watermarking Security: The Effective Key Length

Patrick Bas and Teddy Furon

Abstract

Whereas the embedding distortion, the payload and the robustness of digital watermarking schemes are well understood, the notion of security is still not completely well defined. The approach proposed in the last five years is too theoretical and solely considers the embedding process, which is half of the watermarking scheme. This paper proposes a new measure of watermarking security, called the *effective key length*, which captures the difficulty for the adversary to get access to the watermarking channel. This new methodology is applied here to additive spread spectrum schemes where theoretical and practical computations of the effective key length are proposed. Experimental protocols using either Monte-Carlo simulations, region approximation or rare event probability estimator allow good evaluation of this quantity. For Improved Spread Spectrum (ISS), our analysis exhibits setups where i) the robustness and the security of the scheme are superior to Spread Spectrum and ii) estimating the secret keys from the observations only is not the best way to break the scheme. Moreover a comparison with Correlation Aware Spread Spectrum (CASS) shows that ISS offers a better security than CASS for a given robustness.

Index Terms

Digital Watermarking, Security.

The authors are listed in alphabetical order.

Patrick Bas is with CNRS-LAGIS, Ecole Centrale de Lille, Av. Paul Langevin, 59651 Villeneuve D'Ascq, France. patrick.bas@ec-lille.fr

T. Furon is with Inria research centre Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes, France. teddy.furon@inria.fr

I. INTRODUCTION

From the early beginning of its history, watermarking has been characterized by a trade-off between the embedding distortion and the capacity. The embedding distortion counts how hiding messages degrades the host contents. The capacity is the theoretical amount of hidden data that can be reliably transmitted when facing an attack of a given strength. In practice, the operating point of a watermarking technique is defined by the embedding distortion, the payload, and the robustness. These are well defined and gauged, for instance, by a Document to Watermarking power Ratio DWR, a number of bits per host samples, and a Symbol Error Rate SER at a given Watermark to Noise power Ratio WNR.

Security came as a third feature stemming from applications where these exist attackers willing to circumvent watermarking such as copy and/or copyright protection. The efforts of the pioneering works introducing this new concept first focused on stressing the distinction between security and robustness. An early definition of security was coined by Ton Kalker as *the inability by unauthorized users to have access to the raw watermarking channel* [1].

The problem addressed in this paper is the following: the methodology to assess the security levels of watermarking schemes, proposed in [2], [3], [4], [5], [6], doesn't completely capture T. Kalker's definition. These papers are based on the translation of C. E. Shannon's definition of security for symmetric crypto-systems [7] into watermarking terms. This was the first approach providing important insights on watermarking security. Section II-A presents this past approach in more details and outlines the following fact: this methodology only takes into account the embedding side. How could it capture the '*access to raw watermarking channel*' in Kalker's definition if just half of the scheme is considered? The decoding process plays an important role and this article puts it back into the picture.

This article also challenges this mainstream framework by noticing that watermarking and symmetric cryptography strongly disagree in the following point: In symmetric cryptography, the deciphering key is unique and is the ciphering key. Therefore, inferring this key from the observations (here, say some cipher texts) is the main task of the attacker. The disclosure of this key grants the adversary the access to the crypto-channel. This is implicitly assumed in Shannon's framework. Nevertheless watermarking differs from symmetric cryptography by the fact that several keys can reliably decode hidden messages. Therefore, the precise disclosure of the secret key used at the embedding side is one possible way to get access to the watermarking channel, but it may not be the only one.

As a solution, this article proposes an alternative methodology to assess the security level of a watermarking scheme as detailed in Sect. II-B. It is also inspired by cryptography, but this time, via

the brute force attack scenario. In brief, our approach is based on the probability P that the adversary finds a key that grants him the access to the watermarking channel as wished by Kalker: either a key decoding hidden messages embedded with the true secret key, either a key embedding messages that will be decoded with the true secret key. This gives birth to the concept of *equivalent keys* presented in Sect. III. Our new definition of the security level is called the *effective key length* and is quantified by $\ell = -\log_2(P)$ in bits. This transposes the notion of cryptographic key length to watermarking: the bigger the effective key length, the smaller the probability of finding an equivalent key. This alternative methodology equally takes into account the embedding and the decoding sides. It is also simpler and more practical because the numerical evaluation of the effective key length is made possible (see Sect. V). This is mainly due to the fact that our approach is not based on information theoretical quantities whose closed-form expressions are difficult to derive (if not impossible), and whose estimations by numerical simulation are a daunting task.

The contributions of the paper are the following:

- A new methodology to estimate the security levels of watermarking schemes based on the definition of equivalent keys, the probability of finding such an equivalent key, and its translation in bits (Sect. III).
- The application of this methodology to the Spread Spectrum (SS) and Improved Spread Spectrum (ISS [8]) watermarking schemes under the Known Message Attack[1] (KMA) scenario giving closed-form expressions of the effective key length in Sect. IV.
- An experimental setup in Sect. V for estimating the effective key length, which is applied on SS, ISS, and CASS (Correlation Aware Spread Spectrum [9]).

This paper leads to some evidences shared with the previous information theoretic approach. For instance, SS and ISS schemes have low security levels as soon as the adversary can get some observations under the KMA scenario. However, there are also points of disagreement. For instance, Sect.IV-C2 shades light on classical key estimators revealing that they can be inappropriate to break the scheme. The experimental work leads to significantly different conclusions about the security of CASS than some earlier results.

II. WATERMARKING SECURITY

This section details the previous methodology for evaluating the security levels of watermarking schemes, and then it presents our proposal. The limitations of the two approaches are outlined.

[1]A scenario where the attacker knows the embedded messages of some watermarked contents.

A. The past approach

1) Security is measured by the equivocation: We model the host by a vector $\mathbf{x}$ in set $\mathcal{X}$ extracted from a block of content. Given a secret key $\mathbf{k}$, the embedding modifies this signal into vector $\mathbf{y}$ to hide message m: $\mathbf{y} = e(\mathbf{x}, m, \mathbf{k})$. The secret key is usually a signal: In SS, the secret key is the set of carriers ; in Quantization Index Modulation schemes (QIM), it is the dither randomizing the quantization [10]. This signal is usually generated at the embedding and decoding sides thanks to a pseudo-random generator fed by a seed. However, the attacker has no interest in disclosing this seed, because, by analyzing watermarked contents, it is usually simpler to directly estimate $\mathbf{k}$ without knowing this seed.

The attacker may disclose different kinds of information about the secret key. First, he might get no information at all. This has been qualified as perfect covering in [2] or stego-security in [5]. This happens when there is a total lack of identifiability of the secret key. A partial lack of identifiability stems in different classes of security where the attacker only learns that the secret key lies in a given subset. For instance, in a SS scheme, he may learn that the watermark is added in a given subspace, however he may not identify the secret carriers up to a rotation matrix in this subspace. This is defined as subspace security in [5].

The application of the information theoretic approach of C. E. Shannon allowed to quantify watermarking security levels [2], [6], [3], [4]. This theory regards the signals used at the embedding as random variables (r.v.). Let us denote $\mathbf{K}$ the r.v. associated to the secret key, $\mathcal{K}$ the space of the secret keys, $\mathbf{X}$ the r.v. associated to the host, $\mathcal{X}$ the space of the hosts. Before producing any watermarked content, the designer draws the secret key $\mathbf{k}$ according to a given distribution $p_{\mathbf{K}}$. The adversary knows $\mathcal{K}$ and $p_{\mathbf{K}}$ but he doesn't know the instantiation $\mathbf{k}$. The approach consists in measuring this lack of knowledge by the entropy of the key $H(\mathbf{K}) \triangleq -\int\!\!\!\!\!\!\int_{\mathcal{K}} p_{\mathbf{K}}(\mathbf{k}) \log_2 p_{\mathbf{K}}(\mathbf{k})$ (i.e., an integral if $\mathbf{K}$ is a continuous r.v. or a sum if $\mathbf{K}$ is a discrete r.v.).

Now, suppose the adversary sees N_o observations denoted as $\mathbf{O}^{N_o} = \{\mathbf{O}_1, \ldots, \mathbf{O}_{N_o}\}$. The question is whether this key will remain a secret once the attacker gets the observations. These include at least some watermarked contents which have been produced by the same embedder (same algorithm $e(\cdot)$, same secret key $\mathbf{k}$). These are also regarded as r.v. $\mathbf{Y}$. The observations may also encompass some other data depending on the attack setup (see definitions of WOA, KMA, KOA in [2]). Note that this article focuses on the KMA (Known Message Attack) scenario where observations are pairs of a watermarked content and its embedded message: $\mathbf{O}_i = (\mathbf{Y}_i, M_i)$.

By carefully analyzing these observations, the attacker might deduce some information about the

secret key. The adversary can refine his knowledge about the key by constructing a posteriori distribution $p_{\mathbf{K}}(\mathbf{k}|\mathbf{O}^{N_o})$. The information leakage is given by the mutual information between the secret key and the observations $I(\mathbf{K};\mathbf{O}^{N_o})$, and the equivocation $h_e(N_o) \triangleq H(\mathbf{K}|\mathbf{O}^{N_o})$ determines how this leakage decreases the initial lack of information: $h_e(N_o) = H(\mathbf{K}) - I(\mathbf{K};\mathbf{O}^{N_o})$. The equivocation is always a non increasing function. With this formulation, a perfect covering is tantamount to $I(\mathbf{K};\mathbf{O}^{N_o}) = 0$. Yet, for most of the watermarking schemes, the information leakage is not null. If identifiability is granted, the equivocation about the secret key decreases down to 0 ($\mathbf{K}$ is a discrete r.v.) or $-\infty$ ($\mathbf{K}$ is a continuous r.v.) as the adversary keeps on observing more data.

This methodology needs $p_{\mathbf{X}}$, $p_{\mathbf{K}}$ and $e(\cdot)$ to derive the distribution of the observations and, in the end, the equivocation. There is no use of the decoding algorithm. It has been successfully applied to additive SS and ISS schemes [6], [3] for Gaussian distribution $p_{\mathbf{X}}$ and to the lattice-based DC-QIM (Distortion Compensated Quantization Index Modulation) scheme [4], [11] under the flat host assumption ($p_{\mathbf{X}}$ is constant at the scale of the watermark signal).

2) Limitations: Finding a closed-form expression of the equivocation is a difficult task. As far as the KMA is considered, a closed-form expression is known for SS [6, Eq. (5)]), ISS for $N_o = 1$ [6, Eq. (22)] and CASS for $N_o = 1$ [12, Eq. (6)] (the latter results must be handled with caution as we are not convinced by the Gaussian assumption of the projections) under the Gaussian setup, and for DC-QIM for $N_o = 1$ and $N_o > 1$ for the cubic lattice [4, Eq. (16) and (29)], and asymptotically for some 'good' lattices [4, Eq. (35)], under the flat host assumption. Bounds have been given for ISS with $N_o > 1$ [6, Eq. (22)] and DC-QIM for $N_o > 1$ [4, Eq. (34)]. This latter reference also gives an experimental protocol dedicated to this particular setup in its section III.B. Note also that if there exist practical methods to estimate the mutual information from random sampled vectors [13], these methods are not immune to the curse of dimensionality.

This methodology faces some limited interpretations in practice as well. As written in [6], there is a need to "*translate the equivocation into other measures that result in being more useful for the evaluation of the security from a practical point of view*". Very few papers bridge this gap and only for some particular watermarking schemes: [6, Lemma 1] relates the equivocation to the normalized correlation between the estimated and the true secret keys, which is a quantity of upmost importance for SS watermarking while [4, Eq.(40)] lower bounds the variance of the dither estimation thanks to the equivocation. Even these relations bring little information regarding T. Kalker's basic definition of security, e.g. the ability of the adversary to have access to the watermarking channel. A noticeable exception is the experimental measurements of the Symbol Error Rate (SER) obtained with an estimated secret keys [11, Fig. 6.a].

This last figure is the most similar work w.r.t. to our approach explained below.

B. Our proposal

1) Security is measured by the effective key length: In symmetric cryptography, the security is in direct relationship with the length of the secret key, which is a binary word of L bits. The length of the keys L is the entropy in bits if the keys are uniformly distributed but it is also the maximum number of tests in logarithmic scale of the brute force attack which finds the key by scanning the $|\mathcal{K}|$ potential keys [14]. The stopping condition has little importance. One often assumes that the adversary tests keys until decoded messages are meaningful. We can also rephrase this with probability: If the adversary draws a key uniformly, the probability to pick the secret key is $P = 2^{-L}$, or in logarithmic scale $-\log_2(P) = L$ bits. With the help of some observations, the goal of the cryptanalysts is to find attacks requiring less operations than the brute force attack. A good cryptosystem has a security close to the length of the key L. For instance, the best attack so far on one version of the Advanced Encryption Standard using 128 bits secret key offers a computational complexity of $2^{126.1}$ [15]. Studying security within a probabilistic framework has also been done in other fields of cryptography (for instance, in authentication [16]).

Our idea is to transpose the notion of key length to watermarking thanks to the brute force attack scenario. A naive strategy is to take the size of the seed of the pseudo-random generator as it is the maximum number of tests of a brute force attack scanning all the seeds. Yet, it doesn't take into account how the secret key is derived from the seed, and especially whether two seeds leads to very similar secret keys. Our approach relies on the probabilistic framework explained below, which takes into account that the secret key may not be unique in some sense.

Denote by $\hat{m}$ the message decoded from $\mathbf{y}$ with the secret key $\mathbf{k}$: $\hat{m} = d(\mathbf{y}, \mathbf{k})$. We expect that $\hat{m} = m$, but this might also hold for another decoding key $\mathbf{k}'$. This raises the concept of equivalent keys: for instance, $\mathbf{k}'$ is equivalent to the secret key $\mathbf{k}$ if it grants the decoding of almost all contents watermarked with $\mathbf{k}$ (see mathematical definitions (4) and (5) in Sect. III). This idea was first mentioned in [17], where the authors made the first distinction between the key lengths in cryptography and watermarking.

The fact that the decoding key might not be unique creates a big distinction with cryptography. However, the rationale of the brute force attack still holds. The attacker proposes a test key $\mathbf{k}'$ and we assume there is a genie telling him whether $\mathbf{k}'$ is equivalent to $\mathbf{k}$. In other words, the security of the watermarking scheme relies on the rarity of such keys: The lower the probability P of $\mathbf{k}'$ being equivalent to $\mathbf{k}$, the more secure is the scheme. We propose to define the effective key length as a logarithmic measure of this probability. Note that in our proposal, we must pay attention to the decoding algorithm $d(\cdot)$ because

it is central to the definition of equivalent keys.

Like in the previous methodology, the attack setup (WOA, KMA, KOA) determines the data from which the test key is derived. In this paper, we restrict our attention to the Known Message Attack (KMA - an observation is a pair of a watermarked content and the embedded message: $\mathbf{o}_i = \{\mathbf{y}_i, m_i\}$).

Assessing the security of watermarking within a probabilistic framework is not new. S. Katzenbeisser has also listed the drawbacks of the information theoretic past approach [18]. He especially outlined the lack of assumption on the computing power of the attacker. He then proposed to gauge security as the advantage of the attacker. In a first step, the adversary, modeled by a probabilistic polynomial-time Turing machine, observes contents watermarked with the secret key $\mathbf{k}_1$ or $\mathbf{k}_2$. Then, the designer produces a new piece of content $\mathbf{y}$ and challenges the adversary whether $\mathbf{y}$ has been watermarked with key $\mathbf{k}_1$ or $\mathbf{k}_2$. The advantage is defined as the probability of a right guess minus $1/2$. One clearly sees that a strictly positive advantage implies that the adversary has been able to infer some information about the secret key during the first step. However, the relationship with his ability to access the watermarking channel is not straightforward: the decoding is not considered, and the notion of equivalent keys is missing.

2) Advantages and limitations: The effective key length is simple to be interpreted because it is directly related to Kalker's definition. The fact that we explicitly take into account both the embedder and the decoder may be seen as a limitation or an advantage. It lowers the universality of a study compared to the previous theoretical approach. Yet, it does make sense since our goal is a *practical* security assessment of a specific watermarking scheme. Nevertheless, there is a clear limitation: the analysis holds for a given process inferring a test key from the observations. Since this process is a priori not the best, rigor only allows us to write an upper bound of the effective key length for $N_o > 0$: $\ell \leq -\log(P)$. The same occurs in applied cryptanalysis where the security of a crypto-system holds until a better attack is discovered. Sect. IV-C focuses on this point for the ISS.

III. Definition of the effective key length

This section explains the concept of equivalent keys necessary to define the effective key length.

A. Definition of the Equivalent Keys

We define by $\mathcal{D}_m(k) \subset \mathcal{X}$ the decoding region associated to the message m and for the key $\mathbf{k}$ by:

$$\mathcal{D}_m(\mathbf{k}) \triangleq \{\mathbf{y} \in \mathcal{X} : d(\mathbf{y}, \mathbf{k}) = m\}. \tag{1}$$

The topology and location of this region in $\mathcal{X}$ depends of the decoding algorithm and of $\mathbf{k}$.

To hide message m, the encoder tries to push the host vector $\mathbf{x}$ deep inside $\mathcal{D}_m(\mathbf{k})$, and this creates an embedding region $\mathcal{E}_m(\mathbf{k}) \subseteq \mathcal{X}$:

$$\mathcal{E}_m(\mathbf{k}) \triangleq \{\mathbf{y} \in \mathcal{X} : \exists \mathbf{x} \in \mathcal{X} \text{ s.t. } \mathbf{y} = e(\mathbf{x}, m, \mathbf{k})\}. \tag{2}$$

A watermarking scheme provides robustness by embedding in such a way that the watermarked contents are located far away from the boundary inside the decoding region. If the vector extracted from an attacked content $\mathbf{z} = \mathbf{y} + \mathbf{n}$ goes out of $\mathcal{E}_m(\mathbf{k})$, $\mathbf{z}$ might still be in $\mathcal{D}_m(\mathbf{k})$ and the correct message is decoded. For some watermarking schemes (like QIM without distortion compensation), we have $\mathcal{E}_m(\mathbf{k}) \subseteq \mathcal{D}_m(\mathbf{k})$. Therefore, there might exist another key $\mathbf{k}'$ such that $\mathcal{E}_m(\mathbf{k}) \subseteq \mathcal{D}_m(\mathbf{k}')$. A graphical illustration of this phenomenon is depicted on Fig. 1. In general even if there is no noise, $\mathcal{E}_m(\mathbf{k}) \nsubseteq \mathcal{D}_m(\mathbf{k})$. This means that some decoding errors are made even in the noiseless case. We define the Symbol Error Rate (SER) in the noiseless case as $\eta(0) \triangleq \mathbb{P}\left[d(e(\mathbf{X}, M, \mathbf{k}), \mathbf{k}) \neq M\right]$. Capital letters $\mathbf{X}$ and M explicit the fact that the probability is over two r.v.: the host and the message to be embedded.

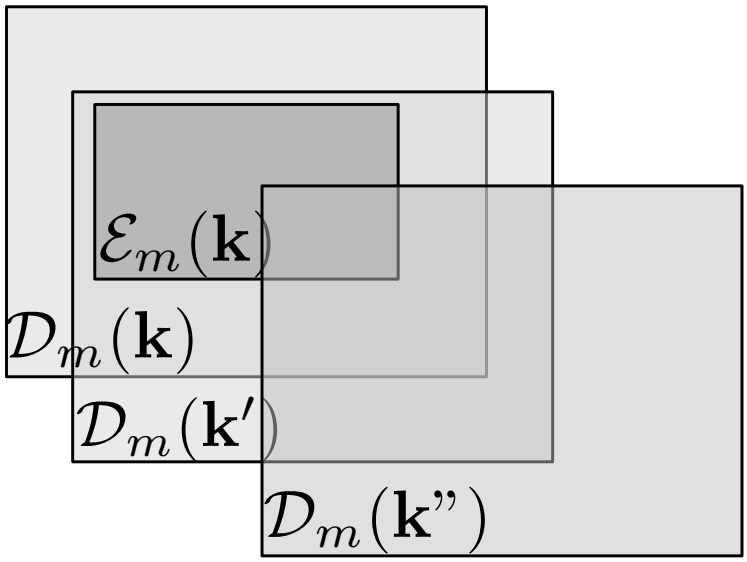

Fig. 1. Graphical representation in $\mathcal{X}$ of three decoding regions. The key $\mathbf{k}'$ belongs the equivalent decoding region $\mathcal{K}_{eq}^{(d)}(\mathbf{k}, 0)$, but not $\mathbf{k}''$.

We now define the equivalent keys and the associated equivalent region. We make the distinction between the equivalent decoding keys (the equivalent decoding region) and the equivalent embedding keys (resp. the equivalent embedding region).

The set of equivalent decoding keys $\mathcal{K}_{eq}^{(d)}(\mathbf{k}, \epsilon) \subset \mathcal{K}$ with $\epsilon \geq 0$ is the set of keys that allows a decoding

of the hidden messages embedded with $\mathbf{k}$ with a probability bigger than $1-\epsilon$:

$$\mathcal{K}_{eq}^{(d)}(\mathbf{k},\epsilon) = \{\mathbf{k}' \in \mathcal{K} : \mathbb{P}\left[d(e(\mathbf{X},M,\mathbf{k}),\mathbf{k}') \neq M\right] \leq \epsilon\}. \tag{3}$$

In the same way, the set of equivalent encoding keys $\mathcal{K}_{eq}^{(e)}(\mathbf{k},\epsilon) \subset \mathcal{K}$ is the set of keys that allow to embed messages which are reliably decoded with key $\mathbf{k}$:

$$\mathcal{K}_{eq}^{(e)}(\mathbf{k},\epsilon) = \{\mathbf{k}' \in \mathcal{K} : \mathbb{P}\left[d(e(\mathbf{X},M,\mathbf{k}'),\mathbf{k}) \neq M\right] \leq \epsilon\}. \tag{4}$$

These sets are not empty for $\epsilon \geq \eta(0)$ since $\mathbf{k}$ is then an element. One expects that, for a sound design, these sets are empty for $\epsilon < \eta(0)$. Note that for $\epsilon = 0$, these two definitions are equivalent to:

$$\mathcal{K}_{eq}^{(d)}(\mathbf{k},0) = \{\mathbf{k}' \in \mathcal{K} : \mathcal{E}_m(\mathbf{k}) \subseteq \mathcal{D}_m(\mathbf{k}')\}, \tag{5}$$

and

$$\mathcal{K}_{eq}^{(e)}(\mathbf{k},0) = \{\mathbf{k}' \in \mathcal{K} : \mathcal{E}_m(\mathbf{k}') \subseteq \mathcal{D}_m(\mathbf{k})\}. \tag{6}$$

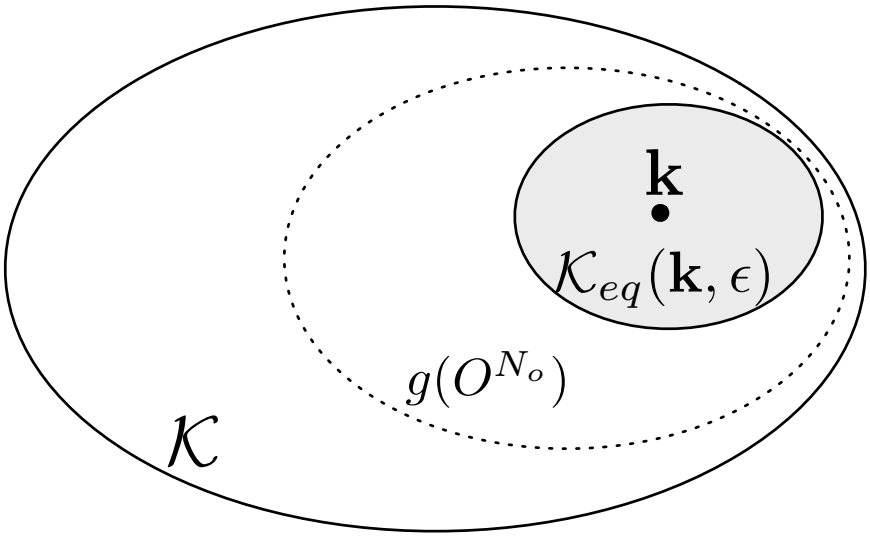

Fig. 2. Graphical representation of $\mathcal{K}$ and the equivalent region $\mathcal{K}_{eq}(\mathbf{k})$. The dotted boundary represents the support of the generative function $g(O^{N_o})$ which is used to draw test keys when the adversary get observations.

B. Definition of the Effective Key Length

The effective key length of a watermarking scheme is now explained using these definitions. For $N_o = 0$, we call $\ell(\epsilon,0)$ the *basic key length*, i.e. the effective key length of a watermarking system when no observation is available. The adversary randomly generates a key and a genie tells him whether this

key gives him access to the watermarking channel. For $\mathbf{K'}$ and $\mathbf{K}$ independent and distributed by $p_{\mathbf{K}}$, the probability of success is:

$$P^{(d)}(\epsilon, 0) \triangleq \mathbb{E}_{\mathbf{K}}[\mathbb{E}_{\mathbf{K'}}[\mathbf{K'} \in \mathcal{K}_{eq}^{(d)}(\mathbf{K}, \epsilon)]] \tag{7}$$

where $\mathbb{E}_{\mathbf{K}}[\cdot]$ denotes the expectation over $\mathbf{K}$.

By analogy with the brute force attack in cryptography, the effective key length translates this probability into bits:

$$\ell^{(d)}(\epsilon, 0) \triangleq -\log_2(P^{(d)}(\epsilon, 0)) \quad \text{bits}, \tag{8}$$

which is also the logarithm of the average number of guesses needed to find an equivalent key.

For $N_o > 0$, the adversary can do a better job by inferring some information about $\mathbf{K}$ from the set of observations $\mathbf{O}^{N_o}$. We denote this inferring process by $\mathbf{K'} = g(\mathbf{O}^{N_o})$. The generative function $g(\cdot)$ is either deterministic (e.g., $\mathbf{k'} = \mathbb{E}[\mathbf{K}|\mathbf{O}^{N_o}]$) or stochastic (e.g., $\mathbf{K'} \sim p_{\mathbf{K}|\mathbf{O}^{N_o}}$). The probability of success is as follows:

$$P^{(d)}(\epsilon, N_o) = \mathbb{E}_{\mathbf{K}}[\mathbb{E}_{\mathbf{O}^{N_o}}[\mathbb{E}_{\mathbf{K'}}[\mathbf{K'} \in \mathcal{K}_{eq}^{(d)}(\mathbf{K}, \epsilon)|\mathbf{O}^{N_o}]]]. \tag{9}$$

Similar definitions are straightforward for $\ell^{(e)}(\epsilon, N_o)$. Note also that for some watermarking schemes like SS or DC-QIM [19], we have $\mathcal{K}_{eq}^{(e)}(\mathbf{k}, \epsilon) = \mathcal{K}_{eq}^{(d)}(\mathbf{k}, \epsilon)$. For the sake of a simple notation, we use $P(\epsilon, N_o)$ and $\ell(\epsilon, N_o)$ in the sequel.

C. Bounds on the Effective Key Length

Because function $g(.)$ might not be optimal in the sense that it doesn't maximize the probability of success, it only gives an upper bound on the effective key length (provided $\epsilon \geq \eta(0)$):

$$\ell(\epsilon, N_o) \leq -\log_2(P(\epsilon, N_o)) \quad \text{bits}. \tag{10}$$

This gives the maximum effort needed by the adversary to break the watermarking system. If the bound is small we can conclude that the security is low. However, if the bound is large, we cannot conclude that the security is high. As for cryptanalysis, the security of the system relies on the state of the art of the attacks, represented here by function $g(\cdot)$.

We conclude this section by stating that the value of the effective key length should be clipped to the size of the seed in bits. The rationale is the following. We assume that the pseudo-random generator is public (Kerckhoff's principle) so that nothing prevents the attacker from using this generator. The worst case for him is when any seed, except the true one, produces a key not in the equivalent set. Then a brute force attack on the seed yields an effective key length of the size of the seed in bits:

$$\ell(\epsilon, N_o) \leq L \quad \text{bits}, \tag{11}$$

which, taking into account (10), leads to:

$$\ell(\epsilon, N_o) \leq \min\left(L, -\log_2(P(\epsilon, N_o))\right) \quad \text{bits}. \tag{12}$$

IV. Mathematical expressions of the effective key length for ISS

The goal of this section is to analyze the trade-off robustness vs. security thanks to our new approach applied to one of the most popular class of watermarking schemes: Improved Spread-Spectrum (ISS) [8].

The embedding is parametrized by (β, γ):

$$\mathbf{y} = e(\mathbf{x}, m, \mathbf{k}) = \mathbf{x} + ((-1)^m \beta - \gamma(\mathbf{x}^\top \mathbf{k}))\mathbf{k}. \tag{13}$$

The host vector $\mathbf{x} \in \mathbb{R}^{N_v}$ is supposed to be a white Gaussian noise of power σ_X^2. The secret key is a vector uniformly drawn on the hypersphere: $\|\mathbf{k}\| = 1$. The embedding distortion is:

$$D = \mathbb{E}[\|\mathbf{y} - \mathbf{x}\|^2] = \beta^2 + \gamma^2 \sigma_X^2 = N_v \sigma_X^2 10^{-\mathsf{DWR}/10}. \tag{14}$$

Thus, fixing the embedding distortion ties the two parameters: $\beta = \sqrt{D - \gamma^2 \sigma_X^2}$ with γ ranging in $[0, \gamma_{\max}]$ and $\gamma_{\max} = \min(1, \sqrt{D}/\sigma_X)$. For $\gamma = 0$, we have the additive SS.

The decoding is given by the sign of the correlation $d = \mathbf{k}^\top \mathbf{y}$. To evaluate the robustness, we assume the addition of an independent white gaussian noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_N^2 \mathbf{I})$. Thanks to a symmetry argument, we only need to consider one message, say $m = 0$. There is a decoding error if $d = \mathbf{k}^\top \mathbf{n} + (1 - \gamma)\mathbf{k}^\top \mathbf{x} + \beta$ is negative. Signals $\mathbf{n}$ and $\mathbf{x}$ being independent, we obtain:

$$\eta(\sigma_N) = \Phi\left(-\sqrt{\frac{D - \gamma^2 \sigma_X^2}{(1-\gamma)^2 \sigma_X^2 + \sigma_N^2}}\right). \tag{15}$$

$\Phi(\cdot)$ being the cumulative distribution function of the Gaussian random variable, the function $\Phi(-\sqrt{x})$ is a decreasing function. As γ increases from 0, this ratio increases (positive derivative) and so is the robustness. This holds for $\gamma < \gamma^{(R)}$, a root of $-\sigma_X^2 \gamma^2 + (\sigma_X^2 + \sigma_N^2 + D)\gamma - D$ equaling [8, Eq.(20)]:

$$\gamma^{(R)} \triangleq \frac{\sigma_X^2 + \sigma_N^2 + D - \sqrt{(\sigma_X^2 + \sigma_N^2 + D)^2 - 4\sigma_X^2 D}}{2\sigma_X^2}. \tag{16}$$

Above this value (if possible), the robustness decreases. ISS is more robust than SS for $0 < \gamma < \bar{\gamma}^{(R)} \triangleq \frac{2}{D + \sigma_X^2 + \sigma_N^2}$.

A. Equivalent region and basic key length

Knowing that the secret keys are on the hypersphere, the attacker picks $\mathbf{k}'$ s.t. $\|\mathbf{k}'\| = 1$ and we denote $\mathbf{k}^\top\mathbf{k}' = \cos(\theta)$. Suppose that $m = 0$ is transmitted. Decoding with $\mathbf{k}'$ yields correlation $d' = \mathbf{x}^\top\mathbf{k}' + (\beta - \gamma.\mathbf{x}^\top\mathbf{k})\cos(\theta)$. From now on, we consider the hyperplane $\mathcal{H} = \text{Span}(\mathbf{k}, \mathbf{k}')$ equipped with the basis $(\mathbf{e}_1, \mathbf{e}_2)$ s.t. $\mathbf{k} = \mathbf{e}_1$ and $\mathbf{k}' = \cos(\theta)\mathbf{e}_1 + \sin(\theta)\mathbf{e}_2$. We denote the projection of $\mathbf{x}$ onto $\mathcal{H}$ by (x_1, x_2). The associated r.v. X_1 and X_2 are i.i.d. and distributed as $\mathcal{N}(0, \sigma_X^2)$. Thus, in the noiseless case, $D' = (1-\gamma)X_1\cos(\theta) + X_2\sin(\theta) + \beta\cos(\theta)$ is Gaussian distributed. A decoding error occurs when $D' < 0$ with probability:

$$\epsilon = \Phi\left(-\frac{\beta\cos\theta}{\sigma_X\sqrt{\sin^2(\theta) + (1-\gamma)^2\cos^2(\theta)}}\right). \tag{17}$$

For $\eta(0) \le \epsilon \le 1/2$, inverting (17) together with (14) shows that $\mathbf{k}'$ is an equivalent key if $\theta \le \theta_\epsilon$ with

$$\cos\theta_\epsilon = \frac{-\Phi^{-1}(\epsilon)10^{\frac{\text{DWR}}{20}}}{\sqrt{N_v + 10^{\frac{\text{DWR}}{10}}(\gamma(2-\gamma)(\Phi^{-1}(\epsilon))^2 - \gamma^2)}}. \tag{18}$$

In words, $\mathcal{K}_{eq}(\mathbf{k}, 0)$ is the hypercap (the intersection of an hypersphere and an hypercone) of axis $\mathbf{k}$ and angle θ_ϵ. $P(\epsilon, 0)$ is the ratio of the solid angles of this hypercap and of the full hypersphere:

$$P(\epsilon, 0) = \left(1 - I_{\cos^2(\theta_\epsilon)}(1/2, (N_v - 1)/2)\right)/2, \tag{19}$$

where I is the regularized incomplete beta function (c.f. Appendix A-A).

B. Trade-off robustness vs. security for $N_o = 0$

For fixed parameters γ, ϵ and DWR, the basic key length is a decreasing function of N_v (see Fig. 5 for SS), contrary to $\eta(\sigma_N)$. This illustrates the trade-off between security and robustness, which is a well known fact for SS in the watermarking security literature [2], [3], [6], [5]. Appendix A gives the asymptotical value of the basic key length:

$$\lim_{N_v\to\infty} P(\epsilon, 0) = \frac{1}{2}\left(1 - \text{erf}\left(\frac{|\Phi^{-1}(\epsilon)|}{\sqrt{2}}10^{\frac{\text{DWR}}{20}}\right)\right). \tag{20}$$

As $N_v \to \infty$, ISS is more robust while its basic key length decreases but does not vanish to 0.

For fixed N_v and DWR, the basic key length ℓ is a decreasing function of $P(\epsilon, 0)$, which in turn is a decreasing function of $\cos^2(\theta_\epsilon)$. Maximizing this latter quantity gives the best security. A simple analysis of the denominator of (18) shows that, when γ increases from 0, ℓ starts from the basic key length of SS and decreases. The basic key length reaches a minimum at $\gamma = \min(\gamma^{(S)}, \gamma_{\max})$, with

$$\gamma^{(S)} \triangleq (\Phi^{-1}(\epsilon))^2/(1 + (\Phi^{-1}(\epsilon))^2). \tag{21}$$

If $\gamma^{(S)} < \gamma_{\max}$, ℓ then increases for $\gamma > \gamma^{(S)}$ and can even be bigger than the basic key length of SS if $\bar{\gamma}^{(S)} < \gamma < \gamma_{\max}$ with $\bar{\gamma}^{(S)} = 2\gamma^{(S)}$. This happens only if $\epsilon > \Phi(-\sqrt{\gamma_{\max}/(2-\gamma_{\max})})$, which equals 0.159 when $\gamma_{\max} = 1$.

The dependence of the basic key length to the embedding parameters is a major difference with the past approach [2], [3], [6]: for a fixed N_v and $N_o = 0$, the security measured by the entropy $H(\mathbf{K})$ is constant and thus independent of γ.

It is interesting to draw the curve giving ℓ as a function of $\eta(\sigma_N)$ when varying γ. Both $\eta(\sigma_N)$ and ℓ decrease as γ starts increasing from 0. Again, this is the trade-off robustness vs. security. Yet, as γ keeps on increasing, the situation gets more complex. If $\gamma^{(R)} < \gamma < \gamma^{(S)}$ (resp. $\gamma^{(R)} > \gamma > \gamma^{(S)}$), then this curve turns on the left (resp. right as in Fig. 7). We enter here a range of γ where security and robustness can both decrease (resp. increase). It can happen that two different γ gives different key lengths while producing the same robustness. This phenomenon has been discovered for ISS in [6, Fig. 4], but for $N_o = 1$. Our new security approach shows that trading security against robustness does not hold in general even for $N_o = 0$. Fig. 7 shows that a clever tuning of γ renders ISS both more robust and more secure than SS, for $N_o = 0$.

C. Effective key length ($N_o > 0$)

This subsection proposes some generative functions $\mathbf{K}' = g(\mathbf{O}^{N_o})$ specific to ISS. First, we try deterministic functions taken as estimators of the secret keys under the KMA scenario. Then, we propose a stochastic function build on these estimators.

1) Estimators of the secret keys: The pirate observes some watermarked vectors $\mathbf{y}_i$ together with their embedded message m_i. We assume that all messages are equal to 0 without loss of generality (otherwise the pirate works with signals $(-1)^{m_i}\mathbf{y}_i$). According to (13), we have $\mathbb{E}[\mathbf{Y}_i] = \beta\mathbf{k}$, and a simple estimator is the empirical average:

$$\hat{\mathbf{k}}_{\text{AVE}} = N_o^{-1} \sum_{i=1}^{N_o} \mathbf{y}_i. \tag{22}$$

This estimator is Gaussian distributed as $\mathcal{N}(\beta\mathbf{k}, \sigma_X^2 N_o^{-1}\mathbf{R})$, with the following covariance matrix:

$$\mathbf{R} = \left(\mathbf{I}_{N_v} + ((1-\gamma)^2 - 1)\mathbf{k}.\mathbf{k}^\top\right). \tag{23}$$

In other words, the variance of this estimator is $\sigma_X^2 N_o^{-1}$ in any direction of the space orthogonal to $\mathbf{k}$, along which this variance is $(1-\gamma)^2$ times smaller. The connection to the appendix A-B with $\mu = \beta$, $\sigma^2 = \sigma_X^2 N_o^{-1}$ and $\sigma_1^2 = (1-\gamma)^2\sigma^2$ provides:

$$P(\epsilon, N_o) \approx \left[1 - \mathcal{F}\left(\frac{N_v - 1}{\tan^2(\theta_\epsilon)(1-\gamma)^2}; 1, N_v - 1, \lambda\right)\right] \Phi\left(\sqrt{\lambda}\right), \tag{24}$$

with

$$\lambda = \frac{\beta^2 N_o}{\sigma_X^2 (1-\gamma)^2} = N_o \frac{N_v 10^{-\frac{\text{DWR}}{10}} - \gamma^2}{(1-\gamma)^2}. \tag{25}$$

Appendix A-B also shows that the effective key length vanishes to 0 as $N_v \to \infty$ whenever $N_o > 0$, which is a strong difference with the basic key length.

The study of the simple average estimator stems into a better idea: we should also take into account that the direction of minimum variance reveals the true secret key. Indeed, the maximum likelihood estimator $\hat{\mathbf{k}}_{\text{ML}} = \arg\max_{\mathbf{k}} \prod_{i=1}^{N_o} p(\mathbf{y}_i|\mathbf{k})$ complies with this idea. Under the Gaussian host assumption and provided $0 < \gamma < 1$, it amounts at minimizing the quantity $\sum_{i=1}^{N_o} (\mathbf{y}_i - \beta\mathbf{k})^\top \mathbf{R}^{-1} (\mathbf{y}_i - \beta\mathbf{k})$. The gradient w.r.t. $\mathbf{k}$ cancels if

$$\mathbf{Y}\mathbf{Y}^\top \hat{\mathbf{k}} = \frac{\beta}{1-(1-\gamma)^2} \mathbf{Y}\mathbf{1}_{N_o}, \tag{26}$$

with $\mathbf{Y}$ the $N_v \times N_o$ matrix $(\mathbf{y}_1, \cdots, \mathbf{y}_{No})$ and $\mathbf{1}_{N_o}$ the $N_o \times 1$ vector composed of ones. As noted in [20], $\mathbf{Y}\mathbf{Y}^\top$ has rank N_o (at most) and is therefore not invertible if $N_o < N_v$. Yet, we circumvent this difficulty by restricting the search to the linear estimator $\hat{\mathbf{k}} = \sum_{i=1}^{No} w_i \mathbf{y}_i = \mathbf{Y}\mathbf{w}$ that maximizes the likelihood. The optimum weights are then $\mathbf{w} = \frac{\beta}{1-(1-\gamma)^2} (\mathbf{Y}^\top \mathbf{Y})^{-1} \mathbf{1}_{N_o}$. To end up with an unbiased estimator, we set

$$\hat{\mathbf{k}}_{\text{ML-LIN}} = \frac{\sum_{i=1}^{No} w_i \mathbf{y}_i}{\sum_{i=1}^{No} w_i}. \tag{27}$$

This brings the property that $\mathbf{k}^\top \hat{\mathbf{K}}_{\text{ML-LIN}} \sim \mathcal{N}(\mu, \sigma_1^2)$ with $\mu = \beta$, $\sigma_1^2 = \sigma^2 (1-\gamma)^2$, and $\sigma^2 = \|\mathbf{w}\|^2 (\sum_i w_i)^{-2} \sigma_X^2$. For β small, N_o small, and γ close to 1, both estimators generate test keys being more often outside the equivalent region than a purely random test key. This leads to the surprise that these estimators, which are deterministic functions of N_o observations, result in an effective key length bigger than the basic key length (i.e. $N_o = 0$). We solve this paradox in the next section.

2) Forging stochastic test keys: We take advantage here of a phenomenon known as *stochastic resonance* in signal processing [21], [22]. We assume that the strategy of the pirate is to first estimate the secret key, and then to artificially randomize it with noise. The estimator is a deterministic function of the observations $\{\mathbf{y}_i\}_{i=1}^{N_o}$, but the test key is a stochastic process due to the noise addition. We set the test key $\mathbf{K}'$ with the following expression:

$$\mathbf{K}' = a\hat{\mathbf{k}} + b\mathbf{N}, \tag{28}$$

where $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_v})$ and $a \geq 0$. We aim here at finding a sound couple (a, b) once the estimator $\hat{\mathbf{k}}$ has been computed and $\mathbf{N}$ has been drawn and takes value $\mathbf{n}$. We impose the constraint $\|\mathbf{k}'\| = 1$:

$$a^2 \|\hat{\mathbf{k}}\|^2 + b^2 \|\mathbf{n}\|^2 + 2ab\hat{\mathbf{k}}^\top \mathbf{n} = 1. \tag{29}$$

The tested key is an equivalent key if it lies in the hypercone of axis $\mathbf{k}$ and angle θ_ϵ which amounts to $\mathbf{k}^\top \mathbf{k}' > \cos(\theta_\epsilon)$. If the estimator is such that $\mathbf{k}^\top \hat{\mathbf{k}}$ is distributed as $\mathcal{N}(\mu, \sigma_1^2)$ (which is the case with the estimators described above) and $\mathbf{N}$ is a white Gaussian noise of unitary variance, then the probability of finding an equivalent key is given by:

$$P = \Phi\left(\frac{a\mu - \cos(\theta_\epsilon)}{\sqrt{a^2\sigma_1^2 + b^2}}\right). \tag{30}$$

The justification of this approach based on stochastic resonance lies in the fact that the couple $(a, b) = (\|\hat{\mathbf{k}}\|^{-1}, 0)$ can produce a negative ratio for small μ appearing in (30): the correlation with $\mathbf{k}$ lies in expectation below the threshold $\cos(\theta_\epsilon)$ yielding to the above-mentioned paradox. In this case, increasing the variance of the projection by the noise addition increases the probability. The same paradox occurs if variance σ_1 is too big (because N_o is small or the estimator is badly designed). In extreme cases, it is better to discard the inaccurate estimator (i.e., $a \approx 0$) and set $b \approx \|n\|^{-1}$. In other words, we are back to the $N_o = 0$ setup. In practice, the pirate finds $(a^\star, b^\star)$ that maximizes the ratio appearing in (30) under the constrainst (29) thanks to a constrained optimization solver.

V. Practical effective key length computations

This section details different experimental protocols to numerically evaluate the effective key length. We first propose a general framework with a high complexity. For correlation-based decoders (such as those of SS, ISS, and CASS), some simplifications occur and stem into a more practical experimental setup. The last subsection shows how to further reduce the complexity with the help of a rare event probability estimator.

A. The general framework

If we are not limited in term of computational power, the probability $P(\epsilon, N_o)$ can be approximated using a classical Monte-Carlo method. We first generate a set of N_1 random secret keys $\{\mathbf{k}_i\}_{i=1}^{N_1}$. For each of them, we also generate N_2 test keys $\{\mathbf{k}'_{i,j}\}_{j=1}^{N_2}$ computed using N_2 distinct sets of N_o observations. Then, an estimation is:

$$\hat{P}^{(d)}(\epsilon, N_o) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} u^{(d)}(\mathbf{k}'_{i,j}, \epsilon), \tag{31}$$

where

$$u^{(d)}(\mathbf{k}'_{i,j}, \epsilon) = \begin{cases} 1, & \text{if} \quad \mathbf{k}'_{i,j} \in \mathcal{K}_{eq}^{(d)}(\mathbf{k}_i, \epsilon) \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

The probability $P^{(e)}(\epsilon, N_o)$ is respectively approximated using the indicator function $u^{(e)}(\cdot)$ of $\mathcal{K}^{(e)}$.

For $N_o = 0$, each test key $\mathbf{k}'_{i,j}$ is independently drawn according to $p_{\mathbf{K}}$. For $N_o > 0$ and a given i, we first generate N_2 sets of N_o observations $\mathbf{O}_j^{N_o}$, $1 \leq j \leq N_2$, depending on $\mathbf{k}_i$, and we resort to a specific process of constructing $\mathbf{k}'_{i,j} = g(\mathbf{O}_j^{N_o})$ (see Sec. III). Secondly, the equivalent region may not have a defined indicator function. In this case, we generate N_t other contents $\{\mathbf{y}_{i,l}\}_{l=1}^{N_t}$ watermarked with $\mathbf{k}_i$ and the test is satisfied if at least $(1-\epsilon)N_t$ contents are correctly decoded using $\mathbf{k}'_{i,j}$. Mathematically, for the decoding equivalence:

$$\mathbf{k}'_{i,j} \in \mathcal{K}_{eq}^{(d)}(\mathbf{k}_i, \epsilon) \overset{\approx}{\Longleftrightarrow} |\{\mathbf{y}_{i,l} \in \mathcal{D}_{m_l}(\mathbf{k}'_{i,j})\}_{l=1}^{N_t}| > (1-\epsilon)N_t. \tag{33}$$

These N_t other watermarked contents play a different role than the N_o watermarked contents used to produce test keys. In this experimental protocol, an estimation of $P^{(d)}(\epsilon, N_o)$ needs $N_1(N_2N_o + N_t)$ embeddings and $N_1N_2N_t$ decodings. Due to the limitation of the Monte-Carlo method, N_1N_2 should be in the order of $1/P^{(d)}(\epsilon, N_o)$ for having a meaningful relative variance of the estimation. The parameter N_t should also be quite big for having a good approximation of the indicator function of $\mathcal{K}_{eq}^{(d)}(\mathbf{k}_i, \epsilon)$. It is reasonable to take $N_t = O(c^{N_v})$ for some constant c where N_v is the dimension of $\mathcal{X}$.

This procedure is generic and it blindly resorts to the embedding and the decoding as black boxes. If we have some knowledge about the watermarking technique, some tricks reduce the complexity of the estimation. First, the probability of finding an equivalent key might not depend on $\mathbf{k}_i$, so that we can restrict to $N_1 = 1$ original key. This is the case for the watermarking techniques studied in this article.

The keystone of our approach is to base the security level on the evaluation of a probability rather than an information theoretic quantity. Nevertheless, the probability to be estimated might be very weak and out of reach of the Monte-Carlo method. Rare event probability estimators such as [23] are more efficient w.r.t. the runtime. Last but not least, if the equivalent set $\mathcal{K}_{eq}^{(d)}(\mathbf{k}, \epsilon)$ is a region described by few parameters, one can directly estimate the parameters instead of using (31). The following subsections put into practice these simplifications.

B. Approximation of the equivalent region $\mathcal{K}_{eq}^{(d)}$ for correlation-based decoding

The equivalent region $\mathcal{K}_{eq}^{(d)}$ depends on the embedding and decoding. For the additive SS, both processes are so simple that we were able to derive closed-form formula of the probability in Sect. IV. We suppose now that the embedding is more complex which prevents theoretical derivations whereas the decoding remains correlation-based. In Sect. VI, CASS [9] plays the role of such an embedding.

For a given host $\mathbf{x}$, we can always express the result of the embedding as

$$\mathbf{y} = e(\mathbf{x}, m, \mathbf{k}) = a(\mathbf{x}, m)\mathbf{k} + b(\mathbf{x}, m)\mathbf{u}_\perp(\mathbf{x}, m), \tag{34}$$

where $\mathbf{k}^\top \mathbf{u}_\perp(\mathbf{x}, m) = 0$. The decoding with $\mathbf{k}'$ is based on the quantity:

$$\mathbf{y}^\top \mathbf{k}' = a(\mathbf{x}, m)\cos(\theta) + b(\mathbf{x}, m).(\mathbf{k}'^\top \mathbf{u}_\perp(\mathbf{x}, m)), \tag{35}$$

whose sign yields the decoded bit $\hat{m}$. It is important to note that the decoding using $\mathbf{k}'$ can be studied in the 2 dimensional space spanned by $(\mathbf{k}, \mathbf{u}_\perp(\mathbf{x}, m))$. The Symbol Error Rate is expressed in term of the CDF of the statistical r.v. $\mathbf{Y}^\top \mathbf{k}'$ which depends on θ, and is thus denoted $\mathsf{SER}(\theta)$. For $\theta = 0$, we have $\mathsf{SER}(0) = \eta(0)$. For $\epsilon \geq \eta(0)$, we define

$$\theta_\epsilon = \max_{\mathsf{SER}(\theta)=\epsilon} \theta. \tag{36}$$

This shows that the equivalent decoding region is a hypercone of axis $\mathbf{k}$ and angle θ_ϵ, which depends on the embedding. The only thing we need is to experimentally estimate angle θ_ϵ. Then, Eq. (19) provides an approximation of the effective key length.

The estimation of θ_ϵ is made under the following rationale. A vector $\mathbf{y}$ watermarked by $\mathbf{k}$ with $m = 0$ is correctly decoded by any $\mathbf{k}'$ s.t. $\mathbf{k}'^\top \mathbf{k} \geq \cos(\theta_\epsilon)$ if its angle ϕ with $\mathbf{k}$ is such that $\phi \in [\theta_\epsilon - \pi/2, \theta_\epsilon + \pi/2]$ (see Fig. 3). In practice, we generate N_t contents $\{\mathbf{y}_i\}_{i=1}^{N_t}$ watermarked with $m = 0$, and we compute their angles $\{\phi_i\}_{i=1}^{N_t}$ with $\mathbf{k}$. Once sorted in increasing order, we iteratively find the angle $\phi_{\min}$ such that $\mathrm{int}((1-\epsilon)N_t)$ vectors have their angle $\phi \in [\phi_{\min} - \pi/2, \phi_{\min} + \pi/2]$ and set $\hat{\theta}_\epsilon = \pi/2 - \phi_{\min}$.

A much lower number N_t of watermarked vectors is needed to accurately estimate one parameter than for a full region of the space. N_t and N_v directly impact the accuracy of $\hat{\theta}_\epsilon$, but since this boils down to the estimation of a single parameter, the order of magnitude of N_t is rather low in comparison with the effective key length. For example, at $N_v = 60$ and $\mathsf{DWR} = 10\,\mathrm{dB}$, we generate $N_t = 10^6$ contents in order to obtain a reliable effective key length of more than 100 bits. Moreover, the angle θ_ϵ is the same for any $\mathbf{k}$ and the estimation is to be done only once. This avoids counting correct decodings over N_t vectors, see (33).

C. Rare event probability estimator

A fast rare event probability estimator[2] is described in [24]. We explain its application for the correlation-based decoder in this article, but we also used it for the DC-QIM watermarking scheme in [19]. This

[2]available as a Matlab toolbox at www.irisa.fr/texmex/people/furon/src.html

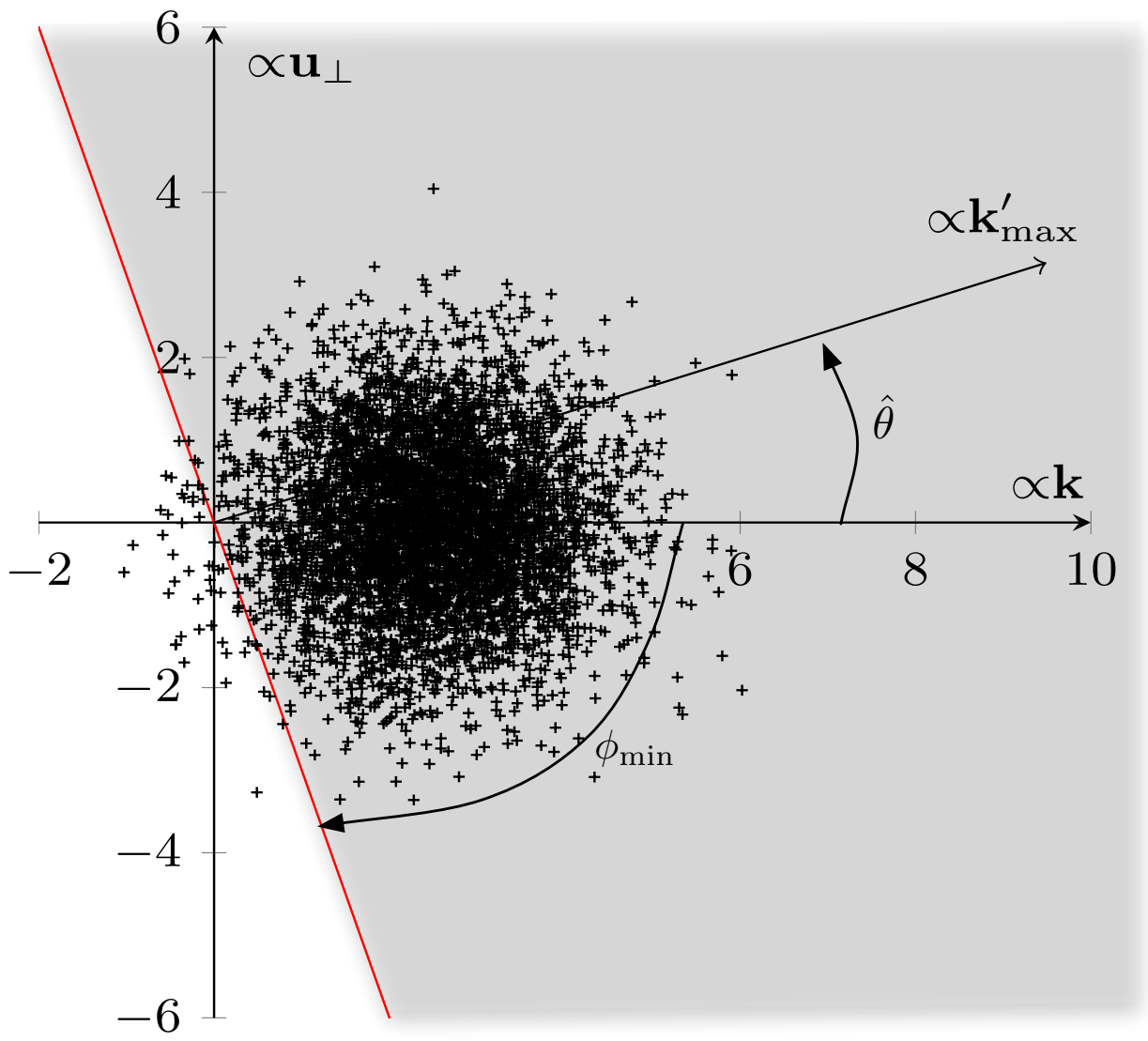

Fig. 3. Projections of $N_t = 5000$ watermarked vectors ($N_v = 60$, $m = 1$) on $\mathbf{k}$ and $\mathbf{u}_\perp$, DWR = 10 dB, $N_v = 60$, $\epsilon = 10^{-2}$. The vector $\mathbf{k}'_{\max}$ correctly decodes $[(1-\epsilon)N_t]$ contents.

algorithm estimates the probability $\mathbb{P}[s(\mathbf{K}') \leq 0]$ under $\mathbf{K}' \sim p_{\mathbf{K}'}$. It needs two ingredients: the generation of test keys distributed according to $p_{\mathbf{K}'}$ and a soft score function $s(\cdot) : \mathcal{K} \to \mathbb{R}$.

For $N_o = 0$, $p_{\mathbf{K}'}$ is indeed $p_{\mathbf{K}}$ from which one can easily sample. In our simulation, we work with an auxiliary random vector $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_v})$. The generator draws $\mathbf{W}$ and outputs a test key $\mathbf{K}' = \mathbf{W}/\|\mathbf{W}\|$. Since the distribution of $\mathbf{W}$ is isotropic, $\mathbf{K}'$ is uniformly distributed over the hypersphere. For $N_o > 0$, $p_{\mathbf{K}'}$ may be unknown. To generate a test key, we first sample N_o independent host contents distributed

as $p_{\mathbf{X}}$, we watermark them with $\mathbf{k}$, and finally apply function $g(.)$. The algorithm draws n such test keys, and iteratively modifies those having a low score. The properties of this algorithm depends on n as given in [24]. Qualitatively, the bigger n is, the more accurate but slower is this estimator.

If $\mathcal{K}_{eq}^{(d)}$ is known (Sect. IV) or approximated (Sect. V-B), the score function is simply a 'distance' between the test key and the equivalent region, this distance being zero if the test key is inside. For the schemes analyzed in this article, we work with $s(\mathbf{k}') = |\cos(\theta_\epsilon) - \mathbf{k}'^\top \mathbf{k}|^+$, with $|x|^+ = x$ if x is positive, and 0 otherwise. If $\mathcal{K}_{eq}^{(d)}$ is not known, we generate N_t contents $\{\mathbf{y}_i\}_{i=1}^{N_t}$ watermarked with $\mathbf{k}$, and the score function is the $\mathrm{int}(\epsilon N_t)$-th smallest 'distance' from these vectors to the set $\mathcal{D}_m(\mathbf{k}')$, where $\mathrm{int}(.)$ denotes the closest integer function. In the end, the algorithm returns an estimation of the probability that $\mathrm{int}((1-\epsilon)N_t)$ of these vectors are correctly decoded when $\mathbf{K}'$ is distributed as $p_{\mathbf{K}'}$.

VI. Results and Discussions

The goal of the experimental part is threefold. First, we wish to assess the soundness of the experimental measurement presented in Sect. V with a comparison to the theoretical results for SS and ISS. Secondly, we aim at analyzing the impact of both the embedding parameters N_v and DWR and the security parameters ϵ and N_o on the effective key length. Third, we would like to illustrate the interplay between security and robustness for both SS, ISS, and CASS (Correlation Aware Spread Spectrum [9]).

The CASS embedding has two parameters (A_1, A_2): $\mathbf{y} = \mathbf{x} + (-1)^m A(m, \mathbf{x}^\top \mathbf{k})\mathbf{k}$, with $A(m, \mathbf{x}^\top \mathbf{k}) = A_1$ if $(-1)^m \mathbf{x}^\top \mathbf{k} > 0$ and A_2 otherwise. The embedding distortion equals $D = (A_1^2 + A_2^2)/2$. For a fixed embedding distortion, $A_2 = \sqrt{2D - A_1^2}$ so that A_1 ranges in $[0, \sqrt{D}]$ while A_2 goes from $\sqrt{2D}$ to $\sqrt{D}$. When $A_1 = A_2 = \sqrt{D}$, CASS is just SS.

The robustness is gauged by the BER $\eta(\sigma_N)$ resulting from an AWGN channel of Watermark to Noise Ratio WNR $= 10\log_{10}(\sigma_W^2/\sigma_N^2)$ dB. As for the security, for $N_o = 0$, we use $N_t = 10^6$ contents to estimate $\hat{\theta}_\epsilon$ as explained in Sect. V-B. The embeddings of SS, ISS, and CASS produce different angles. Then, the rare event probability estimator is used as described in Sect. V-C with $n = 80$.

A. Impact of the parameters N_v, DWR and ϵ

Fig. 4 illustrates the trade-off between security and robustness for SS plotting $\ell(\epsilon, 0)$ as a function of $\eta(\sigma_N)$: for a given N_v, a bigger watermarking *power* implies a bigger robustness but a lower security. However, for a constant watermark *energy* D, the robustness is fixed ($\gamma = 0$ in (15)) while the basic key length increases with N_v.

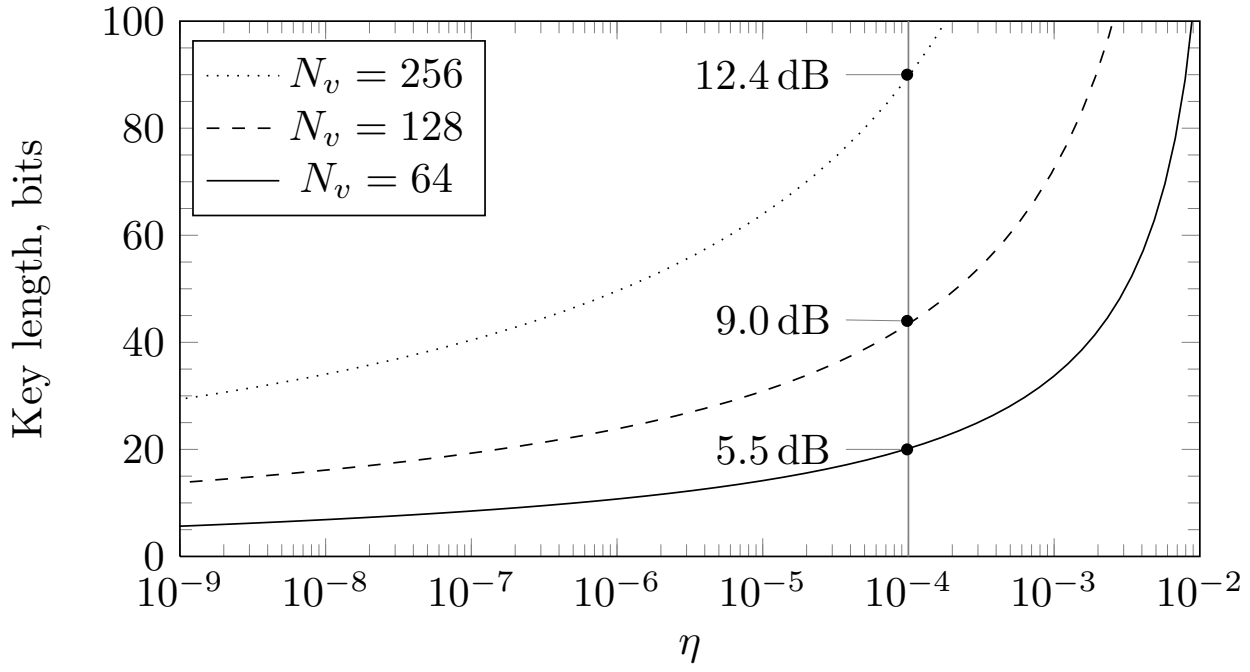

Fig. 4. Security ($\epsilon = 10^{-2}$) vs. robustness (WNR = 0 dB) for SS ($\gamma = 0$). The plot is computed by varying DWR, the ticks show the values of DWR for $\eta = 10^{-4}$.

Fig. 5 shows the evolution of the basic key length w.r.t. N_v for a constant watermark *power* (i.e. fixed DWR). Contrary to a statement of [17, Sec. 4.1], the effective key length is not proportional to N_v. We also note the relatively fast convergence to the strictly positive asymptote (20), especially at high embedding distortions.

Fig. 6 highlights the decrease of this asymptotic effective key length with the watermark power. The basic key length at $\epsilon = 0.01$ is computationally significant, say above 64 bits, only for a DWR greater than 12 dB. We also rediscover the classical sign ambiguity of some security attacks [2] when the adversary ignores the embedded bit: $\lim_{\mathrm{DWR}\to-\infty} \ell = 1$ bit (see Fig. 6 and (20)). The probability of drawing a test key in the adequate half-space tends to 0.5 when the watermarking power is super strong. At last, Fig. 6 shows the decrease of the basic key length with ϵ: the more stringent the access to the watermarking channel, the higher the security is.

B. Interplay between security and robustness

We propose now to analyze the evolution of both the security and the robustness for SS, ISS (γ varying), and CASS (A_1 varying). This is performed under common setups described by WNR, DWR,

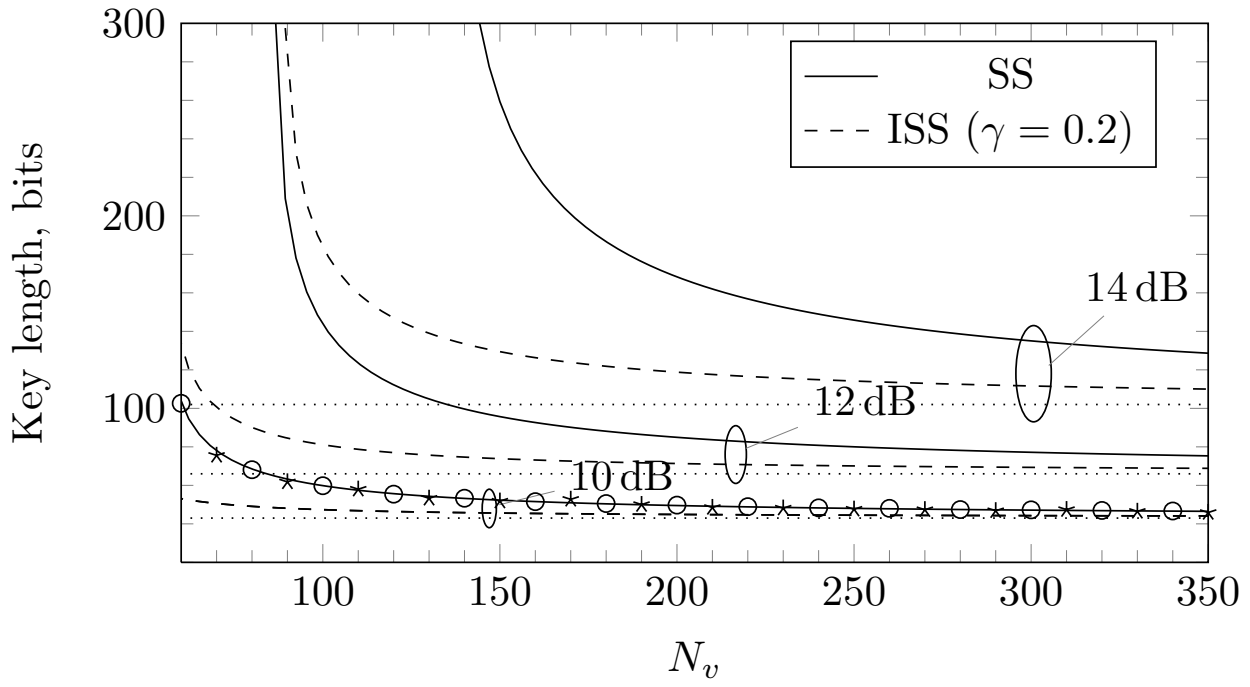

Fig. 5. The basic key lengths for $\epsilon = 10^{-2}$ and $\mathsf{DWR} \in \{10, 12, 14\}$. (plain lines) theoretical expression (19), ($\circ$) estimation of the equivalent region (Sect. V-B) with $N_t = 10^6$, ($\star$) rare event analysis (Sect. V-C) with $N_t = 5.10^4$ and $n = 80$, (dotted lines) the asymptotes (20).

N_v, ϵ, and N_o.

Fig. 7 compares both schemes for $N_v = 180$, $\mathsf{DWR} = 20\,\mathrm{dB}$, $\mathsf{WNR} = -20\,\mathrm{dB}$, $\epsilon = 0.2$, $N_o = 0$. As analyzed in Sect. IV, ISS is more robust than SS for $\gamma \leq \bar{\gamma}^{(R)} \approx 0.95$ and reaches a minimum bit error rate for $\gamma = \gamma^{(R)} \approx 0.55$. The basic key length increases for $\gamma \geq \gamma^{(S)} \approx 0.41$ and ISS is more secure than SS for $\gamma \geq \bar{\gamma}^{(S)}$. This particular setup has a rather high[3] ϵ so that there exists a range $[\bar{\gamma}^{(S)}, \bar{\gamma}^{(R)}]$ where ISS is both more robust and more secure than SS. This convenient representation also shows that CASS is less robust or less secure than ISS under this setup, and that both ISS and CASS can be more robust and more secure than SS.

The setup of Fig. 8 ($N_v = 180$, $\mathsf{DWR} = 14\,\mathrm{dB}$, $\mathsf{WNR} = -7\,\mathrm{dB}$, $\epsilon = 10^{-2}$, $N_o = 0$) shows a more classical behavior already mentioned in [6], [3] where SS is always more secure than ISS, but at the price of a smaller robustness. A surprise is that CASS for $A_1 = 0$ provides basic key lengths more than three times bigger than the one of ISS (at a much bigger $\eta(\sigma_N)$) and that CASS can be more robust and more secure

[3]This scenario is nevertheless relevant if the embedded message went through an error correcting code.

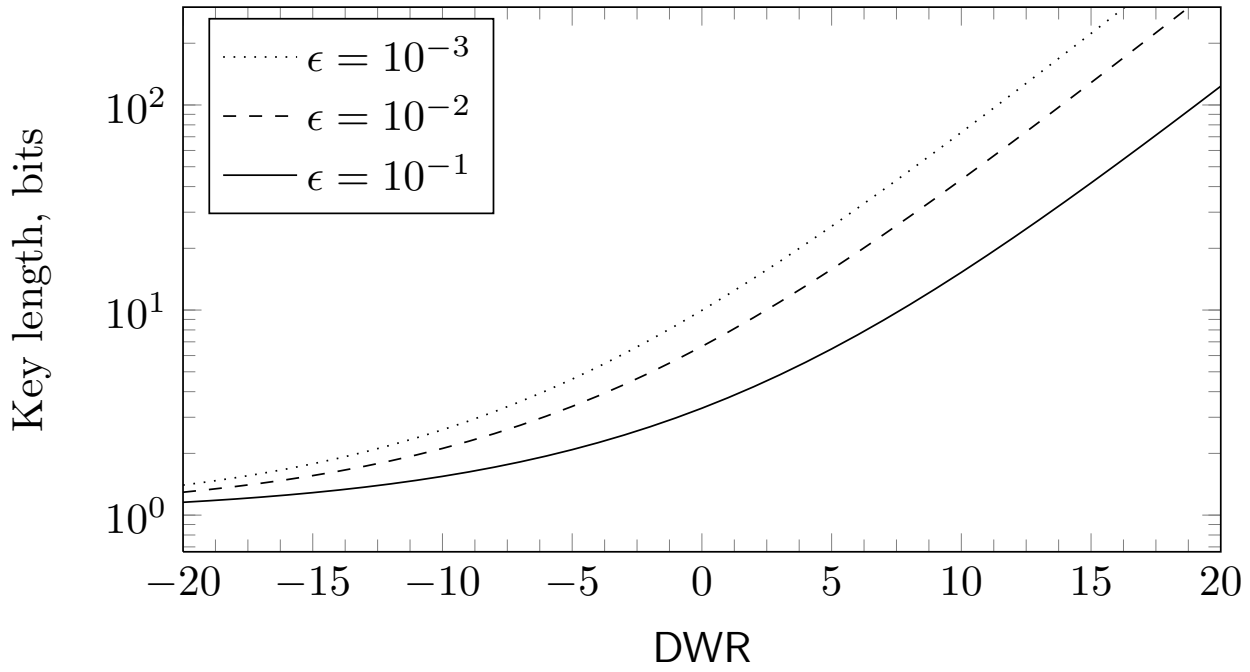

Fig. 6. Basic key length for hosts of infinite length given in (20).

than SS if properly tuned.

C. Impact of N_o

We now evaluate the impact of the number of observations in the KMA setup for SS, ISS, and CASS. For ISS, Fig. 9 illustrates the dramatical collapse of the effective key length when observations are available. For example, at DWR = 10 dB, N_v = 300 and ϵ = 10^{-2}, the effective key length drops from roughly 50 bits to nearly 0 bits within 10 observations. Note also that the approximation (24) is very close to the Monte Carlo estimations. Fig. 10 illustrates the paradox highlighted in Section IV-C1: at DWR = 12 dB, N_v = 256 and ϵ = 10^{-2}, both the average estimator given by (22) and the linear estimator (27) provide upper bounds of the effective key length for $N_o \in \{1, 2\}$ that are above the basic key-length. A classical estimator used so far in watermarking security [20] can lead to a security attack less powerful than a random guess: for N_o = 1 the effective key length obtained using the Average estimator is nearly 3 times larger than the effective key length obtained for N_o = 0! On the other hand, the use of a stochastic estimator minimizing (30) (a numerical solver is needed in this case) yields to what one expects, i.e. a decreasing effective key length w.r.t. N_o. Note also that the key lengths obtained for the stochastic average estimator and the stochastic ML-Lin average estimator are roughly the same.

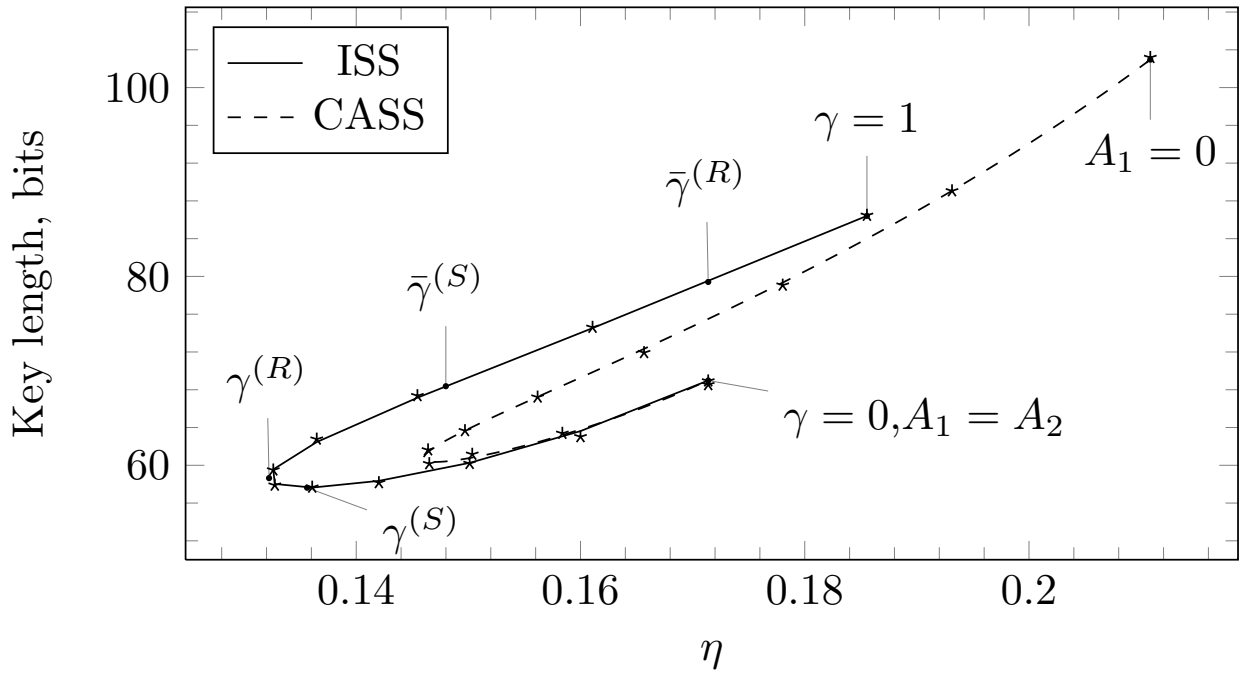

Fig. 7. Robustess and security evolution for ISS and CASS (N_v = 180, DWR = 20 dB, WNR = −20 dB, ϵ = 0.2, N_o = 0). ($\star$) computed using the estimation of the equivalent region (Sect. V-B) with $N_t = 10^6$.

Fig. 11 shows the evolution of the effective key length for CASS for $N_o \in \{0, 1, 10\}$ w.r.t. parameter A_1 for N_v = 180, DWR = 20 dB and $\epsilon = 10^{-2}$. This analysis enables to make several distinctions with the security analysis proposed in [12]:

- for N_o = 1 in the KMA setup, the security of CASS is varying a lot (in the range $[50, 100]$ bits) and can be smaller than the security of SS, which is reached when A_1 takes its maximum value. This contradicts [12, Fig. 2] where the security of CASS is claimed to be almost fixed (a variation of 0.04% of the mutual information), and always slightly bigger than the security of SS.
- we perform a security analysis for any values of N_o whereas the security analysis of [12] was done only for N_o = 1.

D. Validity of the practical approaches

The practical methods (Monte-Carlo, rare-event estimator or equivalent region estimation) match the literal formula (19) and (24) either for small or large effective key lengths on Fig. 5, Fig. 7, Fig. 8, Fig. 9 and Fig. 10. The rare event estimator (Sect. V-C) and the estimator based on $\hat{\theta}_\epsilon$ (Sect. V-B) are

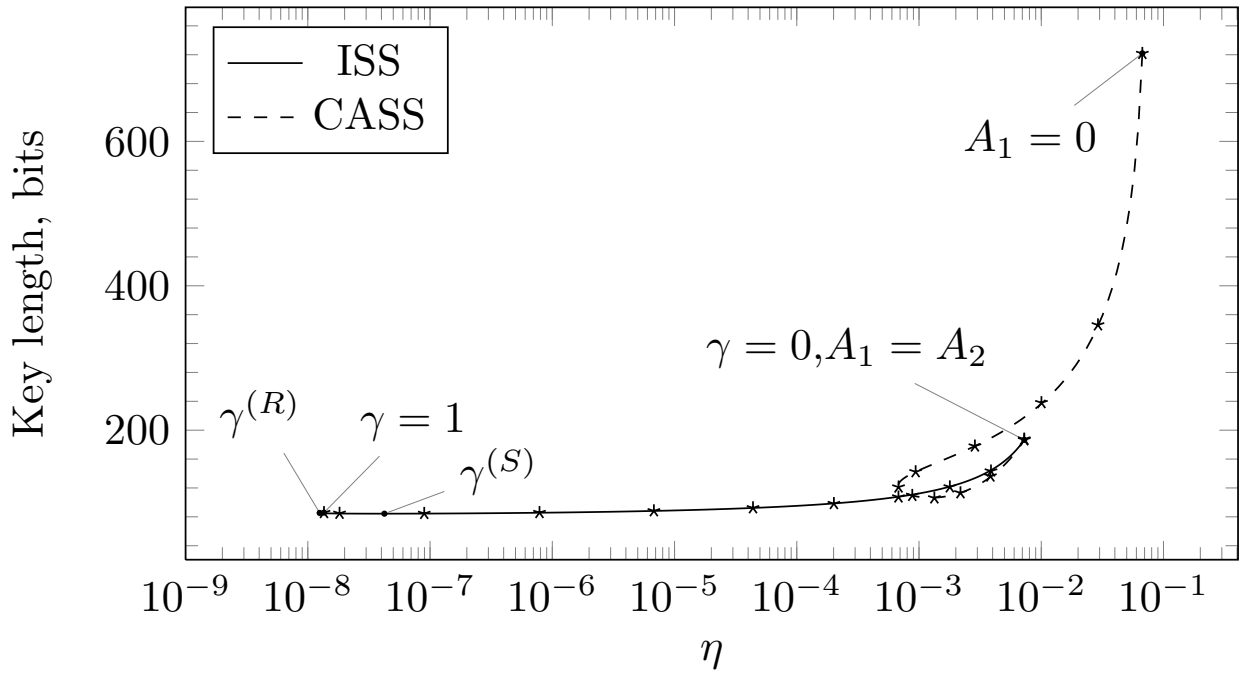

Fig. 8. Robustess and security evolution for ISS and CASS (N_v = 180, DWR = 14 dB, WNR = −7 dB, $\epsilon = 10^{-2}$, N_o = 0). ($\star$) computed using the estimation of the equivalent region (Sect. V-B) with $N_t = 10^6$.

particularly accurate for large key lengths (see Figures 5, 7 and 8), whereas the Monte-Carlo estimator is more efficient for small key length (see Fig. 9).

Note also that it is possible to download the Matlab or R source codes that have been used to draw the different plots of this paper at at http://people.rennes.inria.fr/Teddy.Furon/weckle_toolbox.zip.

VII. Conclusion

This article has proposed a practical measure of watermarking security in direct relationship with the main goal of the adversary: the access to the watermarking channel. The proposed measure is simply the logarithmic scale of the probability of reaching this goal. This is similar to the brute force attack in cryptography, with the notable difference that there might not be a unique key granting the access to the watermarking channel.

We manage to evaluate the effective key length theoretically and/or experimentally for various watermarking schemes such as SS, ISS, and CASS in this paper, and also DC-QIM and zero-bit watermarking schemes in [19], [25]. Combined with a robustness analysis based on the SER, the effective key length is a convenient way to benchmark algorithms from a security/robustness trade-off point of view, at a given

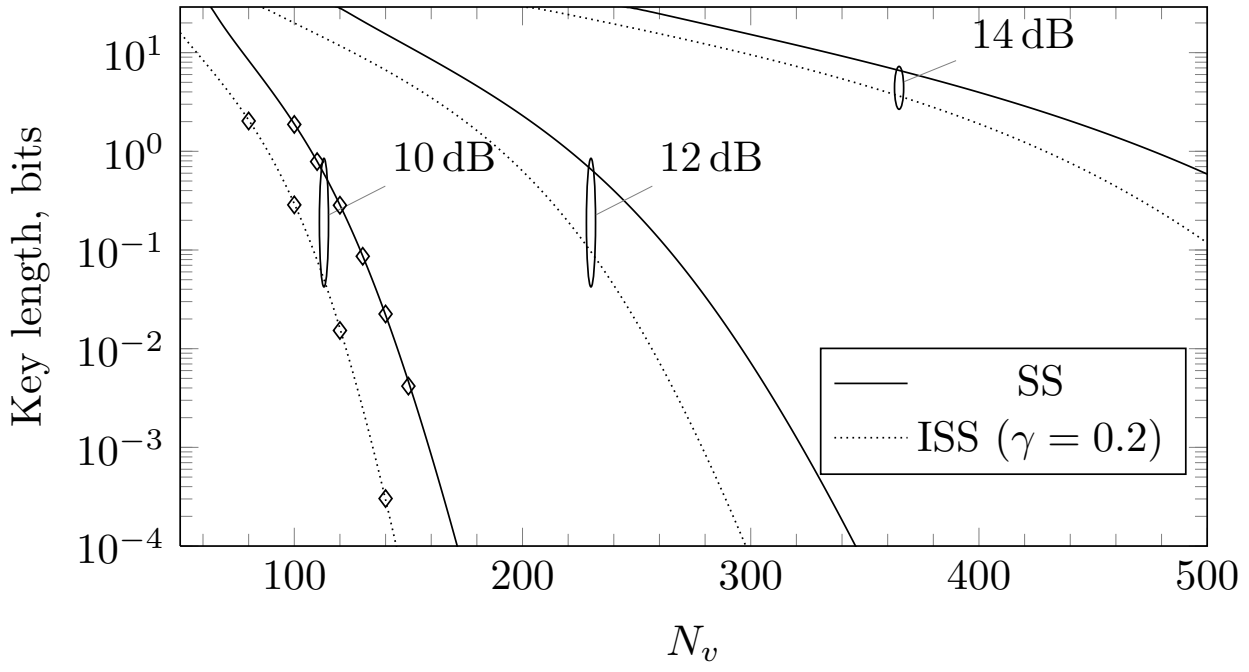

Fig. 9. Effective key lengths for $\epsilon = 10^{-2}$, $N_o = 10$, and different DWR using approximation (24) and Monte-Carlo simulations of Sect. V-A ($\diamond$) with $N_1 = 1$ and $N_2 = 10^6$.

embedding distortion.

When observations are available, for example under the KMA scenario, the adversary has to take care of the way he generates the test keys. We have shown that estimating the secret key might perform worse than a random guess. Our future works include the search of the optimal attack, i.e. minimizing the key length for small N_o.

APPENDIX A
PROBABILITIES FOR IMPROVED SPREAD SPECTRUM

Without loss of generality, we suppose that $\mathbf{k} = \mathbf{e}_1$ and that the attacker uses a test key $\mathbf{K'}$distributed as $\mathcal{N}(\mu\mathbf{e}_1, \Lambda)$ where $\Lambda = \text{diag}(\sigma_1^2, \sigma^2, \ldots, \sigma^2)$. Now, $\mathbf{K'}$ is an equivalent key iff it belongs to the inner single-nappe hypercone, which means $C \triangleq \mathbf{k}^\top \bar{\mathbf{K}}'/\|\bar{\mathbf{K}}'\| \geq \cos(\theta_\epsilon)$. This translates into

$$C = \frac{\sigma_1 U_1 + \mu}{\sqrt{(\sigma_1 U_1 + \mu)^2 + \sigma^2 \sum_{i=2}^{N_v} U_i^2}} \geq \cos(\theta_\epsilon), \tag{37}$$

where $\mathbf{U} = (U_1, \cdots, U_{N_v}) \sim \mathcal{N}(0, \mathbf{I})$. We assume for the moment that $\sigma_1 > 0$.

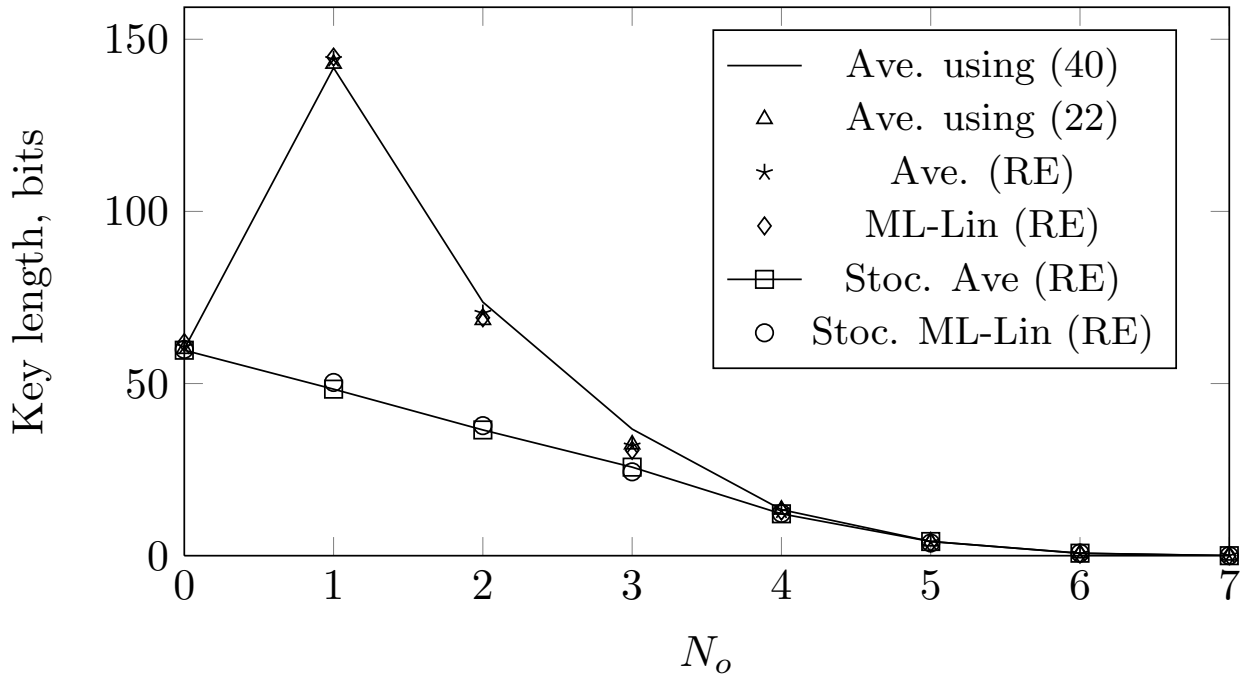

Fig. 10. Evolution of the effective key lengths for ISS using $\gamma = 0.85$ w.r.t. different estimators. "RE" stands for Rare Event probability estimator, "Stoc." the stochastic estimator, "Ave." the average estimator and "ML-Lin" the estimator based of the linear maximum likelihood estimator. $N_v = 256$, DWR = 12 dB and $\epsilon = 10^{-2}$.

A. *Centered F-distribution:* $\mu = 0$ $(N_o = 0)$

A simpler problem is the computation of the probability that a centered Gaussian vector lies inside a two-nappe hypercone. For notation purposes, we set $\tau \triangleq \cos(\theta_\epsilon)$ and $\mu = 0$.

$$\begin{aligned}\mathbb{P}\left[C^2 > \tau^2\right] &= \mathbb{P}\left[\frac{\sigma_1^2 U_1^2}{\sigma_1^2 U_1^2 + \sigma^2 \sum_{i=2}^{N_v} U_i^2} > \tau^2\right] \\ &= \mathbb{P}\left[\frac{U_1^2}{\sum_{i=2}^{N_v} U_i^2} > \frac{\sigma^2}{\sigma_1^2}\frac{\tau^2}{1-\tau^2}\right] \\ &= \mathbb{P}\left[\frac{(N_v-1)U_1^2}{\sum_{i=2}^{N_v} U_i^2} > \frac{\sigma^2}{\sigma_1^2}\frac{(N_v-1)\tau^2}{1-\tau^2}\right]. \end{aligned} \tag{38}$$

We denote $F = \frac{(N_v-1)U_1^2}{\sum_{i=2}^{N_v} U_i^2}$, which is distributed as a Snedecor F-distribution $F(1, N_v-1)$ [26, Sect. 26.6]. Its CDF is given by a regularized incomplete beta function $I_{\frac{x}{x+N_v-1}}(1/2, (N_v-1)/2)$, and

$$\mathbb{P}\left[C^2 > \tau^2\right] = 1 - I_{\frac{\sigma^2\tau^2}{\tau^2(\sigma^2-\sigma_1^2)+\sigma_1^2}}(1/2, (N_v-1)/2). \tag{39}$$

By symmetry around the origin, the probability for the single nappe hypercone is just the half of (39). Eq. (19) is proven by setting $\sigma_1 = \sigma$, i.e. when $\mathbf{K}'$ is a centered and white Gaussian noise.

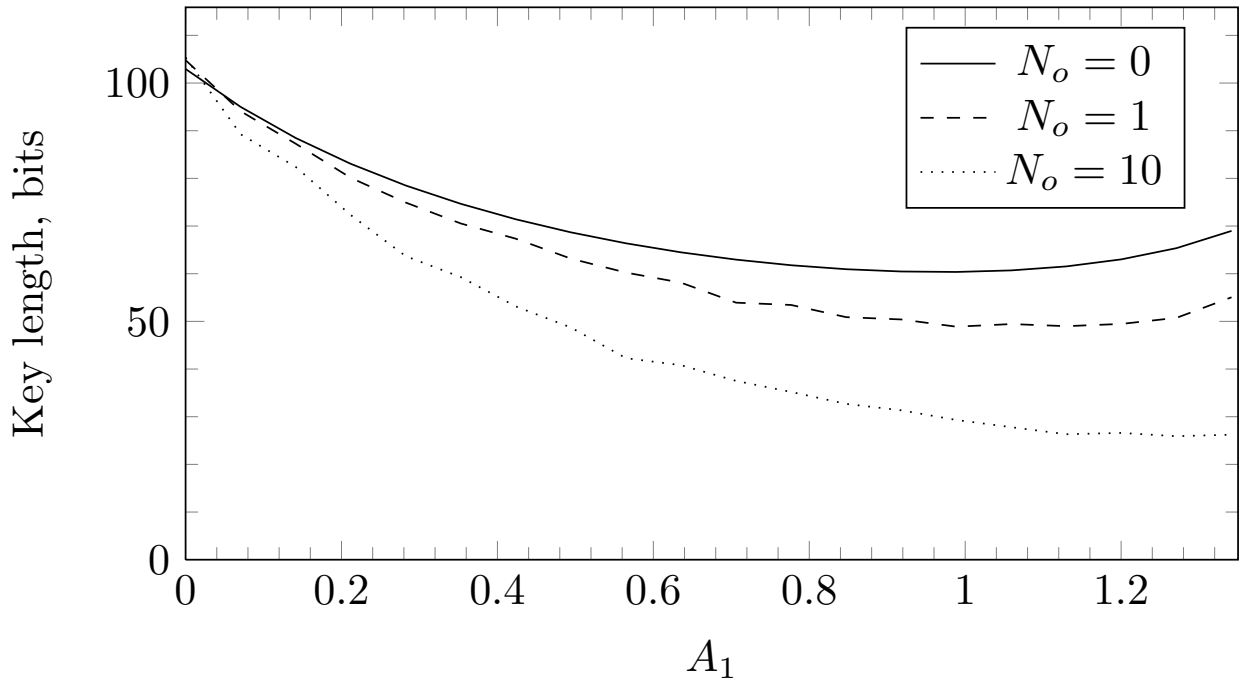

Fig. 11. Evolution of the effective key length of CASS w.r.t N_o and the embedding parameter A_1 (N_v = 180, DWR = 20 dB, $\epsilon = 10^{-2}$). The rare event probability estimator was used.

As $N_v \to \infty$, the distribution of F converges to a χ^2_1 distribution [26, Sect. 26.6.11] while the RHS of the inequality in (38) converges to κ^2 if $\lim_{N_v\to\infty} N_v\tau^2 = \kappa^2$. Therefore, $\lim_{N_v\to\infty} \mathbb{P}[C > \tau] = (1 - \mathrm{erf}(|\kappa|/\sqrt{2}))/2$. This proves (20) thanks to the expression of $\cos(\theta_\epsilon)$ given in (18).

B. non-central F-distribution: $\mu > 0$ $(N_o > 0)$

For $\mu > 0$, F is now $\frac{(N_v-1)(U_1+\sqrt{\lambda})^2}{\sum_{i=2}^{N_v} U_i^2}$, which is a non-central Snedecor r.v. of noncentrality parameter $\lambda = \mu^2/\sigma_1^2$. Its CDF is denoted by $\mathcal{F}(x; 1, N_v - 1, \lambda)$, so that

$$\mathbb{P}\left[C^2 > \tau^2\right] = 1 - \mathcal{F}\left(\frac{\sigma^2}{\sigma_1^2}\frac{(N_v-1)\tau^2}{1-\tau^2}; 1, N_v - 1, \lambda\right). \tag{40}$$

However, the above-mentioned argument of symmetry no longer holds for deriving $\mathbb{P}[C > \tau]$. We propose to write:

$$\begin{aligned} \mathbb{P}[C > \tau] &= \mathbb{P}\left[(C^2 > \tau^2)\&(C > 0)\right] \\ &= \mathbb{P}\left[C^2 > \tau^2 | C > 0\right]\mathbb{P}[C > 0] \approx \mathbb{P}\left[C^2 > \tau^2\right]\mathbb{P}[C > 0], \end{aligned} \tag{41}$$

with $\mathbb{P}[C > 0] = \Phi(\sqrt{\lambda})$. This approximation is accurate for $\lambda \to 0$ and $\lambda \to +\infty$.

F converges to the non-central χ^2_1 r.v. $(U_1 + \sqrt{\lambda})^2$ when $N_v \to \infty$. This makes $\mathbb{P}\left[C^2 > \tau^2\right] \to \mathbb{P}\left[(U_1 + \sqrt{\lambda})^2 > \sigma^2\kappa^2/\sigma_1^2\right]$ if $\lim_{N_v\to\infty} N_v\tau^2 = \kappa^2$. Moreover, if λ increases with N_v as it is the case in (25), we can write as in [27, Proof of Lemma 2.1]:

$$\begin{aligned}\mathbb{P}\left[(U_1 + \sqrt{\lambda})^2 > \kappa^2\right] &= \mathbb{P}\left[U_1^2 + \lambda + 2U_1\sqrt{\lambda} > \kappa^2\right] \\ &= \mathbb{P}\left[U_1 > -\frac{1}{2}\sqrt{\lambda} + \frac{\kappa^2 - U_1^2}{2\sqrt{\lambda}}\right] \overset{\lambda\to\infty}{\to} 1,\end{aligned}$$

and so does $\Phi(\sqrt{\lambda})$. In the end, $\lim_{N_v\to\infty} \mathbb{P}\left[C > \tau\right] = 1$ which shows that the effective key length vanishes to zero as $N_v \to \infty$ provided that $N_o > 0$.

Another way to compute $\mathbb{P}\left[C > \tau\right]$ is by a numerical integration of (41). This amounts to evaluate the following expression:

$$\mathbb{P}\left[C > \tau\right] = \int_0^{+\infty} \Phi\left(\sqrt{\lambda} - \sqrt{x\frac{\sigma^2\tau^2}{\sigma_1^2(1-\tau^2)}}\right) f_{\chi^2}(x; N_v - 1)dx, \tag{42}$$

where $f_{\chi^2}(\cdot, N)$ is the pdf of a χ^2-square with N degrees of freedom.

C. The special case of $\sigma_1 = 0$ $(\gamma = 1)$

The derivations of the appendix so far assumed that $\sigma_1 > 0$. For the special case of ISS with $\gamma = 1$, the host interference is totally cancelled and $\sigma_1 = 0$. For $\mu \geq 0$, we obtain that:

$$\begin{aligned}\mathbb{P}\left[C > \tau\right] &= \mathbb{P}\left[\sum_{i=2}^{N_v} U_i^2 < \frac{\mu^2(1-\tau^2)}{\sigma^2\tau^2}\right] \\ &= P\left(\frac{N_v - 1}{2}, \frac{\mu^2(1-\tau^2)}{2\sigma^2\tau^2}\right),\end{aligned} \tag{43}$$

where $P(k/2, x)$ is the CDF of a chi-squared χ^2_k r.v., i.e., the regularized Gamma function.

REFERENCES

[1] T. Kalker, "Considerations on watermarking security," in *Proc. of IEEE MMSP*, Cannes, France, Oct. 2001, pp. 201–206.

[2] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Trans. Signal Processing*, vol. 53, no. 10, oct 2005.

[3] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Fundamentals of data hiding security and their application to spread-spectrum analysis," in *7th Information Hiding Workshop, IH05*, Barcelona, Spain, June 2005, Lecture Notes in Computer Science, Springer Verlag.

[4] L. Pérez-Freire, F. Pérez-González, T. Furon, and P. Comesaña, "Security of lattice-based data hiding against the Known Message Attack," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 4, pp. 421–439, Dec. 2006.

[5] F. Cayre and P. Bas, "Kerckhoffs-based embedding security classes for WOA data-hiding," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 1, Mar. 2008.

[6] L. Pérez-Freire and F. Pérez-González, "Spread spectrum watermarking security," *IEEE Trans. on Information Forensics and Security*, vol. 4, no. 1, pp. 2–24, Mar. 2009.

[7] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656–715, 1949.

[8] H. S. Malvar and D. A. F. Florêncio, "Improved Spread Spectrum: a new modulation technique for robust watermarking," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, April 2003.

[9] A. Valizadeh and Z. J. Wang, "Correlation-and-bit-aware spread spectrum embedding for data hiding," *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 2, pp. 267 –282, june 2011.

[10] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, Inc., San Francisco, 2001.

[11] L. Pérez-Freire and F. Pérez-González, "Security of lattice-based data hiding against the Watermarked Only Attack," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 4, pp. 593 –610, dec. 2008.

[12] D. Zhang and D.-J. Lee, "Security of CASS data hiding scheme under the scenarios of KMA and WOA," in *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 1797–1800.

[13] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, pp. 066138, 2004.

[14] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*, CRC, 1997.

[15] A. Bogdanov, D. Khovratovich, and C. Rechberger, "Biclique cryptanalysis of the full AES," *ASIACRYPT'11*, 2011.

[16] U. Maurer, "Authentication theory and hypothesis testing," *IEEE Trans. on Information Theory*, vol. 46, no. 4, pp. 1350–1356, July 2000.

[17] I. Cox, G. Doerr, and T. Furon, "Watermarking is not cryptography," in *Proc. Int. Work. on Digital Watermarking*, Jeju island, Korea, Nov. 2006, vol. 4283 of *LNCS*, Springer-Verlag.

[18] S. Katzenbeisser, "Computational security models for digital watermarks," in *Proc. of the Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2005.

[19] T. Furon and P. Bas, "A new measure of watermarking security applied on QIM," in *Proc. of Information Hiding Workshop*, D. Ghosal and M. Kirchner, Eds., Berkeley, CA, USA, May 2012.

[20] L. Pérez-Freire, *Digital Watermarking Security*, Ph.D. thesis, Universidade de Vigo, 2008.

[21] S. Kay, "Can detectability be improved by adding noise?," *Signal Processing Letters, IEEE*, vol. 7, no. 1, pp. 8 –10, jan. 2000.

[22] S. Zozor and P.-O. Amblard, "Stochastic Resonance in Locally Optimal Detectors," *IEEE Transactions on Signal Processing*, vol. 51, pp. no. 12, pp. 3177–3181, 2003.

[23] F. Cérou, P. Del Moral, T. Furon, and A. Guyader, "Sequential Monte Carlo for rare event estimation," *Statistics and Computing*, pp. 1–14, Apr. 2011.

[24] A. Guyader, N. Hengartner, and E Matzner-Lober, "Simulation and estimation of extreme quantiles and extreme probabilities," *Applied Mathematics & Optimization*, vol. 64, pp. 171–196, 2011, 10.1007/s00245-011-9135-z.

[25] P. Bas and T. Furon, "Key length estimation of zero-bit watermarking schemes," in *Proc. EUSIPCO*, Bucharest, Romania, 2012.

[26] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of *National Bureau of Standards Applied Mathematics Series*, Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[27] C. Robert, "On some accurate bounds for the quantiles of a non-central chi squared distribution," *Statistics & Probability Letters*, vol. 10, no. 2, pp. 101 – 106, 1990.

Patrick Bas received the Electrical Engineering degree from the Institut National Polytechnique de Grenoble, France, in 1997 and the Ph.D. degree in Signal and Image processing from Institut National Polytechnique de Grenoble, France, in 2000. From 1997 to 2000, he was a member of the Laboratoire des Images et des Signaux de Grenoble (LIS), France where he worked on still image watermarking. During his post-doctoral activities, he was a member of the Communications and Remote Sensing Laboratory of the Faculty of Engineering at the Université Catholique de Louvain, Belgium. From 2001 to 2009, Patrick Bas worked as a CNRS researcher at Gipsa-Lab, and since 2010 he works at LAGIS, Lille, France. From 2005 to 2008, Patrick Bas was detached from Gipsa-Lab to work as a visiting researcher at the Computer and Information Science Laboratory in the Helsinki University of Technology (Finland). His research interests include synchronisation and security evaluation in watermarking, and steganalysis. From 2004 to 2008, Patrick Bas was co-coordinator of the virtual lab 1 on "watermarking and theory" within the Ecrypt European NoE. Patrick Bas has co-organised the 9th International Workshop on Information Hiding (IH07), the 2nd Edition of the Bows-2 contest on watermarking and the first edition of the BOSS contest on steganalysis.

Teddy Furon received the M.S. degree in digital communications and the Ph.D. degree in signal and image processing from the Ecole Nationale Supérieure des Télécommunications de Paris, Paris, France, in 1998 and 2002, respectively.

From 1998 to 2001, he was a Research Engineer with the Security Lab of Thomson, Rennes, France, working on digital watermarking in the framework of copy protection. He continued working on digital watermarking as a Postdoctoral Researcher at the TELE Lab of the Université Catholique de Louvain, Louvain, Belgium. He also worked in the Security Lab of Technicolor. He is at present a Researcher working within the TEXMEX Team-Project in the Inria Research Center, Rennes, France.

Dr. Furon serves as Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and of the Elsevier Digital Signal Processing Journal. He is an elected member of the IEEE Information Forensics and Security Technical Committee.